

doi:10.19920/j.cnki.jmsc.2023.05.007

# 全景式大数据质量评估指标框架构建研究<sup>①</sup>

安小米<sup>1,2,3</sup>, 黄婕<sup>1,3\*</sup>, 许济沧<sup>1,3</sup>, 王丽丽<sup>1,3</sup>, 洪学海<sup>4,5</sup>, 王志强<sup>6</sup>, 韩新伊<sup>1</sup>  
(1. 中国人民大学信息资源管理学院, 北京 100872; 2. 数据工程与知识工程教育部重点实验室, 北京 100872;  
3. 中国人民大学智慧城市研究中心, 北京 100872; 4. 中国科学院计算技术研究所, 北京 100086;  
5. 中国科学院计算机网络信息中心, 北京 100086; 6. 中国标准化研究院, 北京 100088)

**摘要:** 大数据质量评估工作是促进数字经济、数字社会、数字政府高质量发展的重要保障。本研究针对当前大数据质量评估指标缺少标准化文件来源和全景式评估的相关问题, 梳理出大数据质量评估的多种应用场景, 以综合集成方法论为指导提出由“人理-事理-数理-机理”(HBDA) 构成的全景式大数据质量评估视角。采用内容分析和编码方法, 以大数据质量相关标准化文件丰富指标来源的代表性文献, 经过3名研究人员的两轮筛选, 构建出由56个指标构成的全景式大数据质量评估指标框架。采用案例研究法, 开展B市国际大数据交易所、B市城市管理综合行政执法局和B市大数据中心三个实践案例研究, 有效验证了该框架的正确性和可用性。所提出的HBDA视角下全景式大数据质量评估指标框架, 拓展了全景式PAGE框架在大数据质量评估多场景中的应用; 创新了大数据质量评估多维标准化协同的路径; 对整体提升数字经济、数字社会和数字政府建设中的大数据质量具有战略意义, 对增强数字国家的数据治理能力、大数据驱动的管理与决策能力具有指导意义。

**关键词:** 大数据质量; 评估指标; 框架构建; 全景式框架; HBDA 视角

**中图分类号:** C931; G203 **文献标识码:** A **文章编号:** 1007-9807(2023)05-0138-16

## 0 引言

在新一轮科技革命和产业变革的背景下, 数字经济成为了各国经济增长的主要贡献来源; 数字社会成为了人民生活的新常态; 数字政府成为了各国政府发展建设的主要趋势。作为数字经济、数字社会、数字政府的基础要素, 大数据成为了推动可持续发展、提升数字技术创新能力和促进经济社会高质量发展的关键要素、新动能和新引擎。在国际上, 发达国家早已把数据质量视为数字经济保障并写进法令中。例如, 欧盟GDPR要求数据处理者要按照数据标准保证数据质量; 美国1974年颁布的隐私法包含了关于数据

“准确性、相关性、及时性和完整性”的数据质量要求, 2001年通过了数据质量法(Data Quality Act); 此外, 经合组织(OECD)统计发展战略、加拿大PIPEDA法案均对数据质量做出了明确要求。在我国, 政府大数据质量评估成为个别省市数字政府和大数据发展的重要工作任务。2016年10月25日, 《深圳市促进大数据发展行动计划(2016年—2018年)》明确提出“研究建立政府数据质量评价和管理制度, 规范政府数据采集、共享交换、信息发布等行为, 保障政府数据的质量和可用性。”2019年5月17日, 《贵州省人民政府办公厅关于印发贵州省“一网通办”暂行办法的通知》第二十八条提到, 省政府办公厅对全省“一网通

① 收稿日期: 2021-12-31; 修订日期: 2022-12-26。

基金项目: 国家重点研发计划资助项目(2022YFF0610004); 国家自然科学基金资助项目(72241434; 92046017)。

通讯作者: 黄婕(1996—), 女, 广东茂名, 博士生。Email: huangjie2018@ruc.edu.cn

办”工作定期开展调查评估,调查评估可以委托第三方专业机构,围绕网上政务服务能力、政府数据质量、共享开放程度、云服务质量、网络安全等方面进行。2021年12月,《“十四五”国家信息化规划》指出要“推进数据标准规范体系建设,提高数据质量和规范,建立完善数据管理国家标准体系和数据治理能力评估体系”。大数据质量及其评估工作和标准化研究的重要性愈发凸显。在通过大数据治理提升国家综合竞争力的背景下,国家自然科学基金委员会(NSFC)于2015年9月启动了“大数据驱动的管理与决策研究”重大研究计划,已对大数据管理决策、大数据关联、大数据交易、政府数据开放等重要议题进行了研究<sup>[1]</sup>。大数据质量与价值评估作为大数据资源管理与决策中前沿课题之一<sup>[2]</sup>,与大数据“用”和“造”密切相关,尚有较大研究空间。

当前,数据赋能、数字赋能、数智赋能成为数字化转型的重要内容<sup>[3]</sup>,大数据质量保证是数字经济、数字社会、数字政府高质量发展的前提,大数据质量的评估规则构建已经成为信息化建设、十四五规划和2035年远景目标实现和标准化建设中规则制度建设的重要内容。在“十四五”时期“高质量发展”的主题下,要推动大数据相关产业的高质量发展,加快建立完善大数据质量评估制度。而在实践中,源头数据质量低、质量问题发现和修复困难、数据建设成果评价困难、难以满足监管要求等各种与大数据质量相关的问题依然存在,这些问题影响着大数据共享与利用进程,也给大数据治理带来了困扰,同时当前大数据质量评估的研究仍未得到足够的重视<sup>[4]</sup>。因此,要重视大数据质量评估这一关键环节,通过大数据质量评估工作检查大数据源头质量、检验大数据应用成效、指导大数据工作改进,从而进一步推动高质量和标准化的公共数据共享开放、推动数字基础设施高质量共建共用、提高数字化服务效能,有利于更好地引领数字治理工作高质量发展,推动数字治理流程再造和模式优化,提高决策科学性和服务效率。

从可行性而言,数据质量测度已经在工业数

据、地理信息、汽车工业、软件工程等领域有一定的基础和方案,这些实践为大数据质量评估工作的开展提供了参考。但目前尚未形成统一的大数据质量概念,也缺少大数据质量评估指标框架。本研究拟构建一个全景式大数据质量评估指标框架,面向大数据治理与决策提供大数据质量评估方面的支持,进一步促进大数据资源治理、共享开放和综合利用。

## 1 文献述评

### 1.1 概念界定

大数据具有多种理解和定义,根据国家标准GB/T 35295-2017《信息技术 大数据 术语》,“大数据”是指具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据<sup>[5]</sup>。大数据的定义体现了其具有规模性(volume)、多样性(variety)、高速性(velocity)、多变性(variability)、价值稀疏性(value)等特征。本文采纳该定义并尝试分析大数据特征与大数据质量的关系。

关于大数据质量的定义,目前学界尚未达成统一的认识。根据ISO 25012:2008《软件工程:软件产品质量要求与评估 数据质量模型》,最常用的数据质量定义是“适合使用”<sup>[6]</sup>。数据质量被视为适合数据用户使用的数据标准<sup>[7]</sup>。国家标准GB/T 36344-2018《信息技术 数据质量评价指标》采纳了ISO/IEC 25024:2015中的定义,将数据质量界定为:在指定条件下使用时,数据的特性满足明确的和隐含的要求的程度<sup>[8]</sup>。可见,数据质量本身是一种标准或度量的结果。在一定程度上,“大数据”质量会继承部分“小数据”质量的属性特征<sup>[4,9]</sup>。莫祖英将大数据质量理解为“大数据中适合于进行数据分析、处理、预测等使用过程并满足用户需求的特征”<sup>[10]</sup>。本文将“大数据质量”定义为在多元主体的大数据相关使用活动过程中,大数据能够满足各主体明确和隐含的要求的程度。同时,大数据质量在大数据“5V”特征的共同作用下,相应地也受到数据的规模、来源、时效、

变化和价值密度的影响,大数据质量评估工作要适应大数据规模大、涉及人员多、变化快、价值效用测度的挑战,全方位多角度地评估大数据质量。

## 1.2 相关研究

目前研究主要以数据质量评估为主,关于大数据质量评估的相关文献较少。刘心报等针对新一代信息技术环境下质量管理的新需求、新挑战,提出了全生命周期质量管理体系和全生命周期质量管理方法,其中全生命周期质量评价法对大数据质量的评估具有借鉴意义<sup>[11]</sup>。在目前大数据相关问题的分析视角方面,陈国青等提出了管理决策情境下大数据全景式 PAGE 框架,对大数据粒度缩放、跨界关联和全局视图三大问题特征进行了分析<sup>[12]</sup>。其中,大数据粒度缩放特征的质量问题主要指在不同粒度层级间进行缩放时的数据质

量问题,与数据感知、连接和采集过程、能力相关;跨界关联特征的质量问题主要指在空间外拓和融合时遇到的数据质量问题,与数据治理视角相关;全局视图特征的质量问题主要强调对整体情境下数据质量的管控。

从核心文献中梳理出大数据质量的相关问题,总结出关注大数据质量的相关问题表现和涉及场景,如表 1 所示。从管理过程论和管理要素论的视角出发,梳理出目前大数据质量相关问题涉及到的场景包括:用户服务、安全治理、组织治理、业务融合、数据治理、数据融合、IT 治理、技术融合等八大场景。如果大数据质量得不到保证,“脏数据”会对大数据决策产生误导,不利于大数据有效利用和价值实现,甚至会造成决策失误、成本浪费或经济损失的后果。

表 1 大数据质量相关问题及场景

Table 1 Related issues and scenarios of big data quality

大数据质量问题 分析特征维度	问题描述	相关问题示例	涉及场景
粒度缩放特征维度	指在不同粒度层级间进行缩放时的数据质量问题,与数据感知、连接和采集过程、能力相关	<ul style="list-style-type: none"> <li>● 价值分析质量影响用户服务质量<sup>[13-16]</sup></li> <li>● 数据不一致性<sup>[4,14,17]</sup></li> <li>● 算法对数据分析的准确性<sup>[7,18,19]</sup></li> <li>● 数据并行计算空间和速度<sup>[4]</sup></li> </ul>	<ul style="list-style-type: none"> <li>● 用户服务</li> <li>● 数据治理</li> <li>● 数据融合</li> <li>● IT 治理</li> <li>● 技术融合</li> </ul>
跨界关联特征维度	指在空间外拓和融合时遇到的数据质量问题,与数据治理视角相关	<ul style="list-style-type: none"> <li>● 用户需求自定义<sup>[4,10,15]</sup></li> <li>● 质量报告<sup>[13]</sup></li> <li>● 数据收集问题<sup>[20]</sup></li> <li>● 数据转换问题<sup>[20]</sup></li> <li>● 数据服务扩展性问题<sup>[20]</sup></li> <li>● 数据转变问题<sup>[20]</sup></li> <li>● 大数据提取难度大<sup>[15,26]</sup></li> <li>● 大数据计算困难<sup>[17]</sup></li> <li>● 大数据错误混杂<sup>[17]</sup></li> </ul>	<ul style="list-style-type: none"> <li>● 用户服务</li> <li>● 组织治理</li> <li>● 业务融合</li> <li>● 数据治理</li> <li>● IT 治理</li> <li>● 技术融合</li> </ul>
全局视图特征维度	主要强调对整体情境下数据质量的管控	<ul style="list-style-type: none"> <li>● 用户需求和感知考虑<sup>[4,10]</sup></li> <li>● 数据可信度和保密性考虑<sup>[4]</sup></li> <li>● 大数据质量监控缺失<sup>[13]</sup></li> <li>● 大数据价值密度低<sup>[10,16,19]</sup></li> <li>● 衡量软硬件的综合型质量模型需求<sup>[7]</sup></li> <li>● 缺少知识处理劣质数据查询<sup>[14,17]</sup></li> </ul>	<ul style="list-style-type: none"> <li>● 用户服务</li> <li>● 安全治理</li> <li>● 业务融合</li> <li>● 数据治理</li> <li>● IT 治理</li> <li>● 技术融合</li> </ul>

文献调查结果显示,目前关于大数据质量的研究缺少大数据质量内涵和本质的系统挖掘和全面揭示,也缺少对质量体系诸方面核心思想的提

炼与凝结,基于国家层面和战略层面的技术、人文、管理与大数据相融合的大数据质量研究是重要的研究趋势之一<sup>[9]</sup>。大数据质量是动态的,其



评估工作要与具体的应用目标相结合,而现实研究中往往忽略了大数据质量与大数据价值实现方面关系的研究<sup>[19, 20]</sup>.管理数据质量贯穿大数据生命周期的每个阶段,并进行持续的质量控制和监控,以避免生命周期的所有阶段中的质量失败,大数据质量评价关注的是性能、价值和成本等属性.因此,构建大数据质量评估指标框架不仅要重视大数据的特征,更要关注大数据质量自身的动态性、复杂性和应用目标,更加注重管理、技术、人文思想,在大数据质量管理中贯穿以人为本的用户服务意识,树立注重安全与隐私的安全治理观,通过组织治理和业务融合场景的不同手段提升大数据质量;在数据治理和数据融合场景中提高大数据质量特性,通过 IT 治理和技术融合场景中的新技术、新方法保障大数据质量.

## 2 研究设计

### 2.1 研究目的与研究问题

本文基于大数据质量的重要性和大数据质量评估指标框架及其标准化制度缺乏的现状,通过补充标准化文本研究大数据质量评估指标,构建大数据质量指标框架,为大数据治理工作提供质量保障和标准化路径,为数字要素的治理提供支持.具体研究问题包括:1)构建大数据质量评估指标框架的视角和方法;2)大数据质量评估指标框架的构成要素及关系.

### 2.2 理论基础与分析框架

当数据不仅仅作为被管理的对象存在,通过新技术和方法可以使数据主动赋能时,大数据质量评估既要包括对大数据本身的质量评估,又要体现为大数据赋能过程中对大数据管理过程、大数据新技术与人文环境结合等原则方面的评估.潘云鹤院士认为,人机融合领域是机器人技术发展的重要方向之一,有助于推动我国科技跨越式发展、产业优化升级、生产力整体跃升<sup>[21]</sup>.李国杰院士认为以智能万物为主要特征的人机物三元计算智能技术是推动新经济最有引领性的新技术,通过信息变换优化物理世界的物质运动和能量运动以及人类社会的生产消费活动,提供更高品质的产品和服务,使得生产过程和消费过程更加高

效,更加智能,从而促进经济社会的数字化转型<sup>[22]</sup>.运筹学和系统工程专家顾基发通过“物理事理人理”(WSR)综述综合集成论思想,对复杂系统环境下的大数据治理提供了系统工程、管理科学和人文科学多学科视角的大数据管理决策目标价值实现和方法论的启发<sup>[23]</sup>.可见,大数据与复杂系统环境密不可分,具备技术要素、生产要素和管理要素等特征,对大数据质量的评估应紧紧围绕大数据的要素特征,建立全景式的大数据质量评估指标框架.而在数字化转型背景下,大数据质量评估离不开数字经济、数字社会和数字政府建设的利用需求<sup>[24]</sup>,因此大数据质量评估要特别保障大数据作为数字经济生产要素的“信度”源头治理需求、作为数字社会基础设施要素的“效用”精准治理需求、和作为数字政府业务要素的“尺度”依法治理需求,同时关注反映利益观价值观“温度”的安全治理需求<sup>[25]</sup>.

大数据质量评估场景往往应用于复杂巨系统中.钱学森指出,综合集成方法论是解决开放的复杂巨系统问题的唯一方法论<sup>[26]</sup>.综合集成方法在理论研究层面呈现综合性、整体化的发展趋势,在技术及工程中则形成复杂巨系统工程及实体领导部门.在数字时代背景下,综合集成体系通过发挥大成智慧、协同创新,能极大地落实科学发展观、促进科学创新技术创新和应用创新领域中扩展,以解决人文社会中大数据治理的发展与创新问题<sup>[27]</sup>.本研究以综合集成方法论为依据,围绕大数据质量评估的价值观、业务观、数据观和技术观,充分考虑数字经济、数字社会和数字政府建设中的大数据质量评估需求,创新性地提出“人理-事理-数理-机理”四个视角(Human-Business-Data-Artifact, HBDA 视角),构建有机融合的大数据质量评估指标框架,以评促建,促进社会的高质量发展和智慧城市的可持续发展,促进数字政府、数字经济和数字社会的创新发展.其中,“人理”视角指人的主观意识、素养、价值观对大数据质量的影响,在基于数据赋能的治理基础上,对数据利用的利益观价值观进行评估,关注大数据用户服务和安全治理的“温度”.“事理”视角指业务活动及过程对大数据质量的影响,包括大数据采集、加工、存储、利用及标准化过程中的质量

管控,基于数据依法治理关注大数据组织治理和业务融合的“尺度”。“数理”视角指对数据本身及属性特征进行质量评估,关注数据治理和数据融合的“信度”。“机理”视角指大数据基础设施、设备、环境、算法和技术对质量影响,基于精准治理对数据技术基础设施的评估,关注大数据 IT 治理和技术融合的“效用”。

当前大数据连通了具体丰富的主体、客体、业务活动和场景,使得大数据质量评估具有整体性和交互性。因此,HBDA 视角的提出不仅覆盖了大数据应用中的人类社会和数字空间,同时创新性地数据社会中重视“流动的数据”,促使社会“人理”和物理“机理”在业务“事理”中产生协同效应,发挥“数据”作为资源、资产的价值,在业务活动过程中连通了主体、客体和场景,体现了协同性和自适应性。同时,基于国家战略中对大数据质量评估标准化的重视及目前大数据质量评估研究

中对相关标准的考虑较少这一现状,本研究拟以标准化的思想作为理论补充,具体体现在以标准化文本作为指标的重要来源之一,以弥补当前研究的不足。根据《国家标准化发展纲要》,“标准是经济活动和社会发展的技术支撑,是国家基础性制度的重要方面,标准化在推进国家治理体系和治理能力现代化中发挥着基础性、引领性作用,新时代推动高质量发展、全面建设社会主义现代化国家,迫切需要进一步加强标准化工作”。基于国际标准、国家标准和核心文献提炼相关指标,建立指标架构,采用了标准化的工作思维,是面向国际和人机交互的未来进行架构的质量评估指标框架思路,具有先进性。

根据大数据质量国内外主要标准,为大数据质量的 HBDA 视角提供依据,并对各维度的大数据质量评估指标进行举例和场景说明,构建 HBDA 视角大数据质量评估指标分析框架,如表 2 所示,为大数据质量评估指标编码工作提供总体工作指南。

表 2 HBDA 视角大数据质量评估指标分析框架

Table 2 Big data quality evaluation indicator analysis framework from HBDA perspective

大数据质量评估视角	属性特征分类依据	标准依据	示例	场景
H - 人理	指人的主观意识、素养、价值观对大数据质量的影响,在基于数据赋能的治理基础上,对数据利用的利益观价值观进行评估,关注大数据安全治理的“温度”。	GB/T 36344 - 2018《信息技术数据质量评价指标》 <sup>[8]</sup>	有用性 易于理解性	用户服务 安全治理
B - 事理	指业务活动及过程对大数据质量的影响,包括大数据采集、加工、存储、利用及标准化过程中的质量管控,关注大数据依法治理的“尺度”。	GB/T 36073 - 2018《数据管理能力成熟度评估模型》 <sup>[28]</sup> ISO 8000 - 63《数据质量管理过程测度》 <sup>[29]</sup>	数据质量管理目标 数据质量管理机制	组织治理 业务融合
D - 数理	指对数据本身及属性特征进行质量评估,关注大数据源头治理的“信度”。	GB/T 36344 - 2018《信息技术数据质量评价指标》 <sup>[8]</sup>	规范性 一致性	数据治理 数据融合
A - 机理	指大数据基础设施、设备、环境对质量影响,基于 IT 治理对数据技术基础设施的评估,关注大数据精准治理的“效用”。	GB/T 38673 - 2020《信息技术大数据大数据系统基本要求》 <sup>[30]</sup> ISO/IEC 38500: 2015《信息技术组织 IT 治理》 <sup>[31]</sup> ISO/IEC 25012《软件工程软件产品质量要求和评估数据质量模型》 <sup>[6]</sup> ISO/IEC 25024《系统和软件工程 系统和软件质量要求和评估 数据质量测度》 <sup>[32]</sup>	可恢复性	IT 治理 技术融合

## 2.3 研究方法与技术路线

### 2.3.1 数据来源

在进行大数据质量评估指标来源时,关注大数据属性特征,既要包含国内外关于大数据质量评估的前沿研究,也要覆盖目前指导大数据质量评估的标准化文件。本文的指标数据来源样本为“大数据质量评估”相关的国内外标准和文献,标准数据库检索范围为 ISO<sup>②</sup>、IEC<sup>③</sup>、ITU-T<sup>④</sup> 三大国际标准库和国家标准全文公开系统<sup>⑤</sup>,文献检索范围为 web of science、SCOPUS 数据库和 CNKI 中国知网数据库。此外,研究将大数据质量评估相关领域的信息化和标准化专家推荐的文献作为补充数据来源。根据相关性原则对相关文献进行筛选,共选出“大数据质量”相关的核心文献 26 篇,核心文献覆盖了与大数据质量高度相关的国内外研究和标准,如附录 1 所示。

### 2.3.2 框架构建准则

在筛选完核心文献后,进一步对“大数据质量”核心文献中的指标进行选取。在评估指标框架构建的过程中,需遵循以下原则。

1) 科学性原则。主要体现在研究方案、评估指标和评估方法的科学性。一方面,构建评估指标框架以综合集成论为理论基础,设计科学的研究方案,尽量减少主观因素的干扰。另一方面,评估指标选取兼顾理论层学术研究文献和实践层标准化文件的大数据质量评估指标。

2) 系统性原则。大数据质量评估是多方面的综合评估,也是嵌入到大数据管理过程的评估活动,因此指标框架应该具备系统性。该指标框架应包含大数据质量评估的四个视角,并且呈现出清晰、合理,相互关联的层次结构。为保证评估指标体系具有可信度,在构建时要依据指标间的关联,慎重选择指标并对其进行聚类,尽可能地覆盖评估对象的各个方面,体现出全景式“HBDA”大数据质量评估指标框架的系统性。

3) 可操作性原则。全景式“HBDA”大数据质量评估指标框架最终的落脚点在于使用,因此,构

建的指标体系要具备可操作性,尤其要明晰评估指标的定义,对指标的解释到位,使之清晰易懂,确保操作人员准确把握。

### 2.3.3 指标选取过程及编码有效性保证

通过开放式编码从核心文献中析出指标共 296 个<sup>[7, 8, 10, 14, 18-20, 31, 33-48]</sup>,如附录 2 所示,发现指标主要强调“数理”方面对数据的评估,但也有文献提及对数据管理活动、数据价值观和数据技术条件等方面的评估。

在进行指标概念编码过程中,需要对编码过程进行科学设计和控制,以保障编码的有效性。本研究的编码有效性保证措施主要有:1) 由三名具有定性研究和编码经验的编码人员从“大数据质量”核心文献的评估指标中完成二级指标的编码提取,并进行 Krippendorff Alpha 系数检验;2) 由两名具有定性研究和编码经验的编码人员完成一级指标的聚类 and 编码;3) 所有人员均进行了编码培训,遵循 HBDA 视角大数据质量评估指标分析框架,并随机挑取部分指标进行试编码,对比编码结果进行讨论,完善编码标准达成统一共识;4) 在正式编码时,编码人员独立完成编码工作,在完成编码工作后共同交流编码结果,对意见不一致的地方进行讨论,最后得出采纳指标,保证编码信度。

对指标内容进行分析,归纳出“人理-事理-数理-机理”四个视角,通过 3 名大数据治理研究背景的编码人员独立编码并对每一个指标对其 HBDA 视角进行标注,将不同文献中关于大数据质量相同含义的指标进行合并,得到 167 个二级指标及其类别,如附录 3 所示。上述编码过程的 Krippendorff Alpha 系数为 0.885 6,说明得到的划分结果信度较好。最后,从人理、事理、数理、机理各视角出发,合并名称相近的二级指标成为一级指标,总结同一视角的同一指标的指标描述,得到 HBDA 视角下的 56 个一级指标及其描述,构成全景式大数据质量评估指标框架,如附录 4 所示。

② <https://www.iso.org/obp/ui>

③ <https://www.iec.ch/homepage>

④ <https://www.itu.int/zh/Pages/default.aspx>

⑤ <http://openstd.samr.gov.cn/bz/gk/gb/>



### 2.3.4 研究路线

本研究所采用的研究方法是内容分析法、编码分析方法和案例研究法。在国内外“大数据质量”相关标准和研究的基础上,借鉴 ISO 704:2022 的方法<sup>[49]</sup>,通过内容分析法对大数据质量的核心概念和特征进行提取;基于 HBDA 视角,对大数据质量的评估指标进行编码分析,归纳出二级指标,提炼出一级指标,以构建全景式大数据质量评估指标框架;通过案例研究验证框架和修正指标。

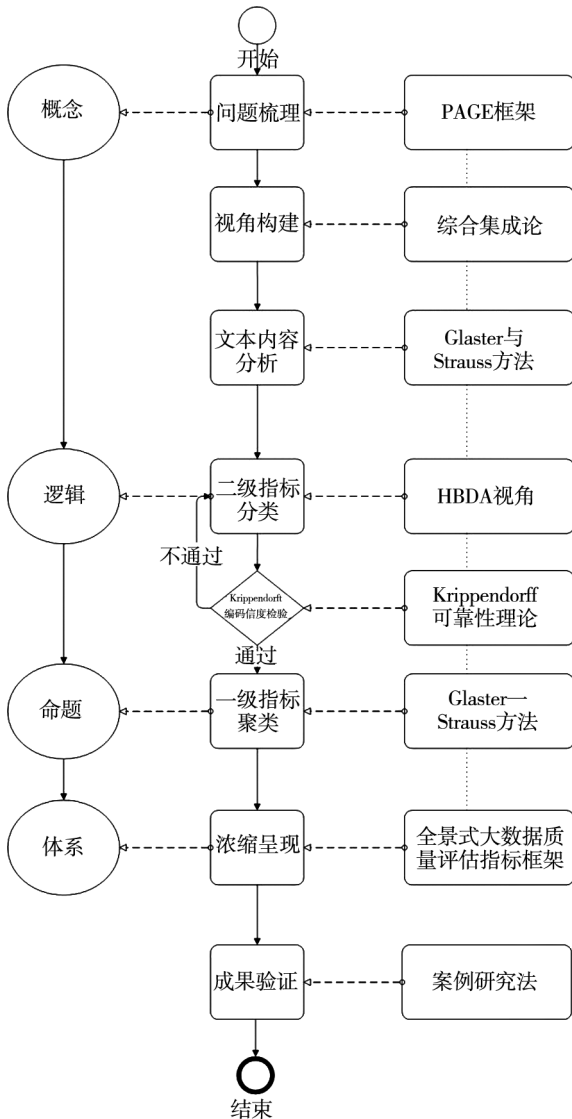


图1 研究技术路线

Fig. 1 Research technical route

本研究的研究路线分为以下步骤:第一步,调查大数据质量国内外标准与文献,根据大数据管理与决策全景式 PAGE 框架的三大问题特征,对文献中大数据质量相关问题进行梳理,总结出大

数据质量相关问题出现的八大场景;第二步,运用综合集成方法论,以大数据质量相关标准为依据,从大数据质量评估的价值观、业务观、数据观和技术观中归纳出 HBDA 视角,提出 HBDA 视角大数据质量评估指标分析框架;第三步,对大数据质量国内外标准与文献进行文本内容分析,对文献中的大数据质量评估指标进行开放式编码,提取核心概念和内涵特征;第四步,基于 HBDA 视角,初步聚类二级指标,并通过 Krippendorff Alpha 系数信度检验;第五步,进一步研究和分析大数据质量评估指标之间的关系,进行聚类得到一级指标,构成全景式大数据质量评估指标框架;第六步,全景式大数据质量评估指标框架的案例研究,通过数字经济、数字社会、数字政府领域的实践案例研究对全景式大数据质量评估指标框架的正确性和可用性进行验证。

## 3 研究过程与发现

### 3.1 全景式大数据质量评估指标框架的构成要素及关系

数据质量全面提升的 HBDA 全景式大数据质量评估框架与全景式 PAGE 框架产生呼应,人理、事理、数理、机理与理论范式(P)研究方向相关联,强调各个视角要素的多元共治和大数据质量评估;数理和机理与分析技术(A)研究方向相关联,强调基于数据的大数据质量分析方法和支撑技术;事理、数理、机理与资源治理(G)研究方向相关联,既关注大数据质量的机制设计评估,也包括大数据、大数据技术资源的协同管理;人理和机理与使能创新(E)研究方向相关联,重点关注了大数据高质量带来的价值创造与模式创新,以及大数据通过特有的技术优势赋能质量保障和创新。在要素内容方面,该框架的构成要素可直接映射到大数据三大问题特征的质量相关问题,也可对大数据质量的相关工作进行指导,具体来说:

“人理-粒度缩放”评估一级指标包括可理解性和可操作性,强调从用户视角关注大数据质量的数据化;“人理-跨界关联”评估一级指标包括从业人员状况、可访问性和相关性,强调用户与大数据治理环境中其他要素的相互作用;“人理

- 全局视图”评估一级指标包括价值性和可用性,强调大数据质量管控过程中对用户的价值观和利益观的考虑.

“事理-粒度缩放”评估一级指标包括浏览次数、浏览下载比、融合性、时段下载比和实时性,强调大数据质量管理活动指标的数据化指标;“事理-跨界关联”评估一级指标包括大数据交换与集成、共享性和可衔接性,强调大数据管理活动中与其他要素相关联的指标;“事理-全局视图”评估一级指标包括安全性、大数据标准体系、法律法规数量和管理制度,强调大数据业务活动中影响全局的标准、规章、制度等指标.

“数理-粒度缩放”评估一级指标包括:数据固有质量、规范性、及时性、解释性、精确性、外观质量、真实性、可靠性、客观性、冗余性、数据到达率、数据规模、数据专业性、透明性、完整性、唯一

性、无错误、信息量、一致性、语法质量、置信度、准确性,强调数据本身的可以在不同属性之间进行缩放粒度的指标;“数理-跨界关联”评估一级指标包括多样性和数据清洗粒度,强调数据要素与其他要素相关联的评估指标;“数理-全局视图”评估一级指标包括合法性、全面性、有效性和增值性,强调数据要素在全局视域中产生影响的指标.

“机理-粒度缩放”评估一级指标包括硬件可靠性、硬件可扩展性、存储利用实时性和系统易用性,强调数据技术基础设施的粒度指标;“机理-跨界关联”评估一级指标包括合理性和可维护性,强调大数据技术环境及其和算法与其他要素的关联性指标;“机理-全局视图”评估一级指标包括安全性、数据分析质量、数据建模质量、效率和数据预测的准确性,强调机理要素能够对大数据质量全局产生影响的指标.

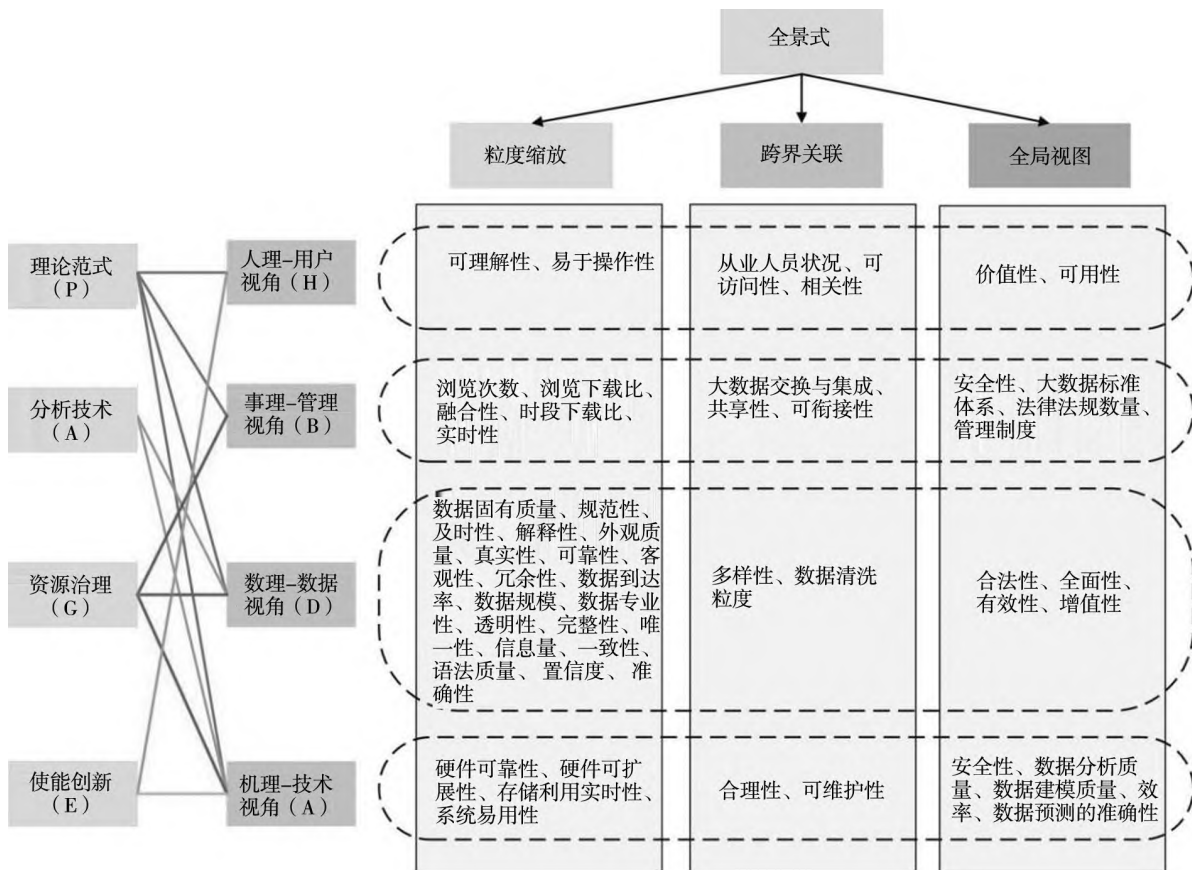


图2 全景式大数据质量评估指标框架

Fig. 2 Panoramic big data quality evaluation indicators framework

### 3.2 全景式大数据质量评估指标框架案例研究

#### 3.2.1 案例研究设计

案例研究是社会科学研究的典型方法之一.

当需要回答“是什么”“怎么样”“为什么”的问题,或研究者几乎无法控制研究对象,或关注的重心是当前现实生活中的实际问题时,适合采用案



例研究法开展研究<sup>[50]</sup>. 为了进一步探讨所构建的全景式“HBDA”大数据质量评估指标框架的正确性和可用性,本研究对数字经济、数字社会、数字政府三大领域下的大数据质量评估进行深入探索,开展案例研究.

在案例样本选择方面,本研究遵循了相关性、针对性、必要性、典型性和便利性原则,考虑了数字经济、数字社会和数字政府、数字政府领域的可用性验证,同时考虑了全景式大数据质量评估框架在实践中的正确性和有效性验证,选取了B市国际大数据交易所、B市城市管理综合行政执法局和B市大数据中心的大数据质量管理评估应用开展案例研究.

在案例数据收集方面,研究团队以三个机构的网络调查资料(包括来自官方网站的政策文件、新闻报道等)、田野调查资料(2021年12月9日在B市大数据中心开展实地调研,2021年12月22日在B市城市管理综合行政执法局开展实地调研)、座谈交流资料(2022年5月23日与B市国际大数据交易所负责人员开展座谈交流)为主要数据来源.

### 3.2.2 数字经济领域:B市国际大数据交易所大数据质量评估案例分析

B市国际大数据交易所(简称“B数所”)<sup>⑥</sup>于2021年3月31日成立,是政府与企业、社会企业之间、境内与境外的数据流通桥梁和服务平台,是中国首家基于“数据可用不可见,用途可控可计量”新型交易范式的数据交易所. B数所的高质量数据开放运营带来了巨大的社会效益,大大提高了小微企业办理相关业务的效率,间接支撑了实体经济,在数字经济领域具有典型代表性.

从事理-管理视角(B)来看,B数所在全局视图方面表现出色:1)在大数据标准体系方面,《B数据交易服务指南》作为B数所的重要标准规范文件和特色规则保障,在大数据交易的数据管理、交易机制、交易环节、跨境数据流动和交易安全管理都做了详细的规定.《B数据交易服务指南》引用了GB/T 25069-2010《信息安全技术 术

语》、GB/T 36344-2018《信息技术 数据质量评价指标》、GB/T 36343-2018《信息技术 数据交易平台 交易数据描述》、GB/T 37728-2019《信息技术 数据交易平台 通用功能要求》、GB/T 37932-2019《信息安全技术 数据交易服务安全要求》和GB/T 35273-2020《信息安全技术 个人信息安全规范作为交易数据质量保障》作为重要的规范参考,体现了大数据质量标准化协同;2)在管理制度上,划定了参与主体的职责,国际大数据交易的交易供方和其他衍生服务供方均需要保障数据的质量,或提供数据质量认证,平台方、运营机构在数据治理的过程中对交易主体和数据质量进行认证,明确规范和标准;3)在安全性管理上,严格遵守交易安全可控原则,确保交易主体可信赖,数据质量可靠,交易目的特定明确,数据流通使用可控,交易记录可查,并搭建数据安全事件预警和响应机制,确保数据安全和交易安全.

为了保障作为公共产品属性的数据的质量评估中立性、权威性和可信性,B数所作为数据交易供方,在B市金融公共数据专区聘请了专门的机构进行数据质量评估,定期发布《B市金融公共数据专区数据质量评价报告》. 数据质量评估机构作为数据中介服务体系建设的一部分,在大数据质量评估和保障中发挥了重要作用.

从数理-数据视角(D)来看,B数所其中一类重要数据是“经过质量认证的数据”,需要保障数据的真实性、完整性、准确性和有效性,确保数据可信、可用. 例如,B市金融公共数据专区已汇聚金融机构开展信贷业务所“亟需、特需”的工商、司法、税务等多维数据25亿余条,覆盖14个部门机构、240余万市场主体,公共数据汇聚质量和更新效率均处于全国领先水平,实现全国首个公共数据授权运营模式落地,而专门的评估机构对金融公共数据专区的评估指标包括“准确性、一致性、完整性、规范性、时效性和可访问性”.

从机理-技术视角(A)来看,B数所曾推出“数字交易合约”,利用区块链技术整合数字身份、价值标定、溯源追踪等能力,利用多方安全计

<sup>⑥</sup> <https://www.bjindex.com/>

算技术、隐私加密技术等,为数据主体签发证书,在数据确权登记、访问、分析、计算、交易过程中,将完整操作过程上链存储,保障数据的来源可追溯、内容防篡改、主权可确认、利益可分配,大大提升了大数据质量。

### 3.2.3 数字社会领域:B市城市管理综合行政执法局大数据质量评估案例分析

B市城市管理综合行政执法局<sup>⑦</sup>基于城市管理与综合执法数字化转型构建了符合中国城市特点的城市治理体系,以城市管理综合执法大数据平台建设为契机,再造大数据条件下的综合执法流程并重塑基层城市治理新格局,创新了B市数字化综合执法与智慧城市治理新模式。

从人理-用户视角(H)来看,B市城市管理综合行政执法局注重与市民开展合作,通过“市民城管通”APP获得可及、可得的社会数据。

从事理-管理视角(B)来看,B市城市管理综合行政执法局自2021年1月起,参考国标、行标、地标、其他委办局的标准、百度地图等产品的标准,制定了核心元数据、数据目录标准、数据元素标准、代码集、数据交换规范、资源目录、数据标识规范和数据仓库标准,但缺少参考标准映射表和通用术语标准。数据质量的评估反馈机制应与业务结合。作为数据使用方,市城管执法局在使用数据过程中对数据质量核对校验,为保障感知设备采集的社会数据的质量,市城管执法局需要制定数据采集业务过程的数字基础设施的数据质量标准,并将标准转换为机器可读的程序,植入到感知设备中。此外,市城管执法局不仅依靠巡查录入的业务模式,实现数据的持续更新,保障数据质量的准确性、时效性,还应将数据质量的评估反馈机制与业务结合,建立包含“支撑业务率”的指标体系,评估从市大数据平台中获取的数据的质量,并反馈质量问题到相应责任部门。

从数理-数据视角(D)来看,B市城市管理综合行政执法局的大数据多样性体现在:历史执法数据、热线举报数据、专项执法工作数据、物联网感知数据及社会共享数据。在可衔接性上,通过数

字化流程规范与再造,物联网传感器保证数据可追溯。

从机理-技术视角(A)来看,B市城市管理综合行政执法局重点打造“一库一图一网一端”(综合执法信息库、执法动态图、执法协同网和移动执法终端APP)的市区街一体化综合执法大数据平台,实现整合、分析、服务、监管、指挥五项功能。

### 3.2.4 数字政府领域:B市大数据中心大数据质量评估案例分析

B市大数据中心是B市经信局下的公益性事业单位。为了服务于领导辅助决策,B市大数据中心基于政府、社会侧等不同来源的数据,开展了数据治理和加工分析。有效解决了数据更新不及时、字段缺失、非结构化数据转化难等数据质量低下问题,保障了数据可信有用。

从事理-管理视角(B)来看,B市大数据中心侧重于大数据质量管理活动和管理制度方面的管理。B市大数据中心作为数据平台方,其数据由B市各政府部门归集至B市大数据平台。在大数据交换与集成方面,实施数据质量全过程监控,具体表现为政府数据在B市大数据平台进行流转时,其更新变化被全程留痕,形成数据血缘。系统的全过程质量监控机制保障了动态数据的可追溯性和在静态时点上的可信性,并记录了不同业务系统数据质量的提升过程。在共享性和可衔接性方面,针对具有质量问题的原始数据,B市大数据平台不仅保留原始数据,还在原始数据下,形成“具体质量问题说明、质量评价规则及修改说明”,并反馈给数据形成方(业务部门),由数据形成方进行修改。在大数据标准体系和法律法规方面,B市大数据中心主要以《政务数据汇聚共享规范》(DB11/T 1919-2021)为依据制定关于数据质量稽核规则,《B市大数据和云计算发展行动计划(2016年—2020年)》等规划政策中提及规范使用数据的重要性,但其法律法规的数量和范围仍有待丰富。在管理制度方面,B市大数据中心开展数据质量检测,对问题数据进行三副本存储,生

<sup>⑦</sup> <http://cgi.beijing.gov.cn/col/col3224/index.html>

成经过多重检验的数据质量评估报告,系统实时更新各委办局业务系统的数据质量排名并纳入绩效考核体系;建立了月报季评等机制和评估反馈机制,以此激励委办局进行数据质量管理。

从数理-数据视角(D)来看,B市大数据中心关注了粒度缩放、跨界关联和全局视图三方面的问题。《政务数据汇聚共享规范》(DB11/T 1919-2021)对业务部门数据的可读性、完整性、准确性、一致性、时效性做出了具体要求,数据形成方负责数据形成、汇聚共享过程的数据质量,B市大数据中心在数据一致性、及时性、完整性、规范性等方面给予客观评价考核。在数据清洗方面,B市大数据平台依据可定制的数据质量稽核规则,对汇聚后的各业务系统的数据进行数据清洗,检测识别出质量问题,形成质量报告。B市大数据平台通过目录区块链系统将原始数据标准化处理后,通过融合分析形成主题层,将高质量的数据资源根据主题进行分类后,提供给数据利用者,从而提高大数据的有效性和增值性。

从机理-技术视角(A)来看,B市大数据中心在技术互操作性上采用全集约方式,在大数据平台上提供数据资源、工具和整体性解决方案。B市大数据中心具有上链数据的质量监督权,在技术方面,利用区块链技术和数据质量检测识别追踪等技术全过程监督数据质量,并将由多方多重检验的数据质量评估报告传输到业务系统,再由数据形成方下载质量评估报告,从源头修正数据质量问题。在硬件可靠性、硬件可扩展性和系统易用性上,2021年,随着B市大数据行动计划向智慧城市2.0建设的纵深发展,在领导驾驶舱基础上有序开展“B智”规划建设,把“B智”与“B办”“B通”和“七通一平”作为B市智慧城市建设的基础设施一盘棋部署、一体化推进;《2022年市政府工作报告重点任务清单》中明确提出要“持续推进‘B智’功能完善和推广使用”“深化‘一网慧治’,采用领导驾驶舱架构,面向不同层级用户提供云服务”。

### 3.3 全景式大数据质量评估框架验证

案例研究表明,HBDA视角下的全景式大数据质量评估框架在以下方面具有正确性和可用性:1)大数据质量管理问题的发现;2)大数据质

量管理成功经验的总结;3)大数据质量管理标准化协同有效路径的揭示4)大数据质量管理工作的持续改进。

#### 3.3.1 大数据质量管理问题的发现

通过全景式大数据质量评估发现,当前数字经济、数字社会和数字政府建设中的大数据质量管理在HBDA视角分析下存在一些问题。具体表现为:1)人理-用户视角(H)下,在大数据利用实际需求方面对用户的可理解性、价值性、可用性方面的关注较少,在从业人员的素质和能力和意识方面缺乏明确的要求;2)事理-管理视角(B)下,大数据标准存在但体系不够完善,大数据标准协同机制尚未确立,当同一业务涉及多家单位数据时,出现多家单位内部制定的数据标准不同,统计数据口径不一致的问题;3)数理-数据视角(D)下,存在数据准确性、真实性等质量问题;4)机理-技术视角(A)下,数据分析质量、数据建模质量、数据预测的准确性没有得到足够的关注。

#### 3.3.2 大数据质量管理成功经验的总结

三个案例中的大数据质量管理评估结果在PAGE框架方面提供了先进经验和相关启示:在理论范式(P)方面,鼓励从HBDA各个视角实现多元共治保障大数据质量的方式;在分析技术(A)方面,从“数理”和“机理”视角出发,积极开发大数据质量分析和保障技术,促进大数据平台技术创新,通过区块链、联邦计算、隐私计算等算法创新应用,促进数据融合和技术融合,实现数据治理和IT治理协同推进;在资源治理(G)方面,从“事理”、“数理”和“机理”视角出发,建立健全大数据质量评估机制,包括质量评估分析机制、反馈机制、绩效机制等,整合大数据资源有效利用的人力、数据和技术资源,形成协同化的资源治理格局;在使能创新(E)方面,从“人理”和“机理”视角出发,以人为中心,更加重视大数据的可理解性和价值性等,不断创新大数据释放价值的模式,同时推动技术变革。

#### 3.3.3 大数据质量管理标准化协同有效路径的揭示

在以上三个案例中,B市大数据中心是B数所和B市数字城管的数据提供方之一,在实际的



数据利用过程中,B市数字城管会反哺B市大数据中心,将大数据质量报告返回给大数据中心.而B数所则会委托第三方数据质量评估机构开展数据质量和价值评估.三个机构的大数据质量评估结果表明,在HBDA视角下,不再仅仅关注数据本身的质量,更加关注了大数据质量在用户视角、管理视角和技术视角方面的保障要素,丰富了复杂性环境下的数据治理,关注了跨层级、跨领域、跨场景、跨平台的大数据质量协同治理问题.

#### 3.3.4 大数据质量管理工作的持续改进

根据案例评估结果,全景式的大数据质量管理在将来,应更加重视数字经济、数字社会和数字政府领域的应用需求,进一步推进大数据价值释放、提升大数据质量向纵向深发展;在人理-用户视角方面,充分重视大数据运营、大数据应用服务、大数据治理等方面的利益相关者需求,加强用户视角的考虑;在事理-管理视角方面,通过大数据的浏览量、下载量、融合性等微观指标侧面反映大数据质量,通过大数据交换、共享与集成等优化过程质量评估;在数理-数据视角方面,更加注重数据本身的增值性考虑;在机理-技术视角,对技术应用的数据分析质量、数据预测的准确性等整体效益指标进行关注.

## 4 结束语

### 4.1 研究结论

本研究在大数据管理与决策背景下,提出了HBDA视角下全景式大数据质量的评估框架.与已有的数据质量评估框架相比,该框架为大数据

质量改进提供了“人理-事理-数理-机理”多维度标准化协同的路径,对满足用户服务、安全治理、组织治理、业务融合、数据治理、数据融合、IT治理、技术融合等多种应用场景下的大数据质量评估需求提供了大数据质量全景式、全方位的评估指标及框架.此外,数字经济、数字社会、数字政府建设中的大数据质量评估案例研究揭示:该框架在大数据质量管理问题的发现、大数据质量管理成功经验的总结、大数据质量管理标准化协同有效路径的揭示和大数据质量管理工作的持续改进四方面具有广阔的应用前景.

### 4.2 未来研究方向及应用前景

本研究对HBDA视角下的大数据质量评估多维标准化路径发现和全景式PAGE框架在大数据质量评估多场景中的应用验证,在案例样本数量和案例场景类型方面待进一步丰富和完善.

未来研究将持续改进所提出的HBDA视角下全景式大数据质量评估指标框架,丰富全景式PAGE框架在大数据质量评估更多场景中的应用;发现大数据质量评估多维标准化协同的更多路径及规律;明晰其对数字经济、数字社会和数字政府建设中的大数据质量整体提升的政策指引和制度规则要求,提出其对增强数字国家的数据治理能力、大数据驱动的管理与决策能力的技术合规要求.研究应用前景广阔,对统筹数字经济、数字社会和数字政府建设中的大数据质量管理政策和标准具有战略意义,对建立健全大数据质量评估制度和标准体系、通过保障大数据质量增强数字国家的数据治理能力和大数据驱动的管理与决策能力具有现实意义.

### 参考文献:

- [1]陈国青,曾大军,卫强,等.大数据环境下的决策范式转变与使能创新[J].管理世界,2020,36(2):95-105+220.  
Chen Guoqing, Zeng Dajun, Wei Qiang, et al. Transitions of decision-making paradigms and enabled innovations in the context of big data[J]. Journal of Management World, 2020, 36(2): 95-105+220. (in Chinese)
- [2]徐宗本,冯芷艳,郭迅华,等.大数据驱动的管理与决策前沿课题[J].管理世界,2014,254(11):158-163.  
Xu Zongben, Feng Zhiyan, Guo Xunhua, et al. Leading-edge issues of big data driven management and decision[J]. Journal of Management World, 2014, 254(11): 158-163. (in Chinese)

- [3]陈国青,任明,卫强,等.数智赋能:信息系统研究的新跃迁[J].管理世界,2022,38(1):180-196.  
Chen Guoqing, Ren Ming, Wei Qiang, et al. Data-intelligence empowerment: A new leap of information systems research [J]. Journal of Management World, 2022, 38(1): 180-196. (in Chinese)
- [4]Ramasamy A, Chowdhury S. Big data quality dimensions: A systematic literature review[J]. JISTEM-Journal of Information Systems and Technology Management, 2020, 17: e202017003.
- [5]全国信息技术标准化技术委员会.信息技术-大数据-术语:GB/T 35295-2017[S].北京:中国标准出版社,2017:1.  
China National Information Technology Standardization Technical Committee. Information technology-Big data-Terminology: GB/T 35295-2017[S]. Beijing: Standards Press of China, 2017: 1. (in Chinese)
- [6]Software engineering: Software product Quality Requirements and Evaluation (SQuaRE): Data quality model; ISO/IEC 25012: 2008[S/OL]. [2021-10-21]. <https://www.iso.org/standard/35736.html>.
- [7]Mohammad A, Mohammad M, Ahmad A, et al. Big data quality factors, frameworks and challenges[J]. Compusoft, 2020, 9(8): 3785-3790.
- [8]全国信息技术标准化技术委员会.信息技术-数据质量评价指标:GB/T 36344-2018[S].北京:中国标准出版社,2018:2-5.  
China National Information Technology Standardization Technical Committee. Information technology: Evaluation indicators for data quality: GB/T 36344-2018[S]. Beijing: Standards Press of China, 2018: 2-5. (in Chinese)
- [9]刘冰,庞琳.国内外大数据质量研究述评[J].情报学报,2019,38(2):217-226.  
Liu Bing, Pang Lin. Review of domestic and international research on big data quality[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(2): 217-226. (in Chinese)
- [10]莫祖英.大数据质量测度模型构建[J].情报理论与实践,2018,41(3):11-15.  
Mo Zuying. Construction of big data quality measurement model[J]. Information Studies: Theory & Application, 2018, 41(3): 11-15. (in Chinese)
- [11]刘心报,胡俊迎,陆少军,等.新一代信息技术环境下的全生命周期质量管理[J].管理科学学报,2022,25(7):2-11.  
Liu Xinbao, Hu Junying, Lu Shaojun, et al. The entire life cycle quality management in the new generation of information technology environment[J]. Journal of Management Sciences in China, 2022, 25(7): 2-11. (in Chinese)
- [12]陈国青,吴刚,顾远东,等.管理决策情境下大数据驱动的研究和应用挑战——范式转变与研究方向[J].管理科学学报,2018,21(7):1-10.  
Chen Guoqing, Wu Gang, Gu Yuandong, et al. The challenges for big data driven research and applications in the context of managerial decision-making: Paradigm shift and research directions [J]. Journal of Management Sciences in China, 2018, 21(7): 1-10. (in Chinese)
- [13]Taleb I, Serhani M A, Bouhaddioui C, et al. Big data quality framework: A holistic approach to continuous quality management[J]. Journal of Big Data, 2021, 8(1): 76.
- [14]Lee D. Big data quality assurance through data traceability: A case study of the national standard reference data program of Korea[J]. IEEE Access, 2019, (99): 1.
- [15]Immonen A, Pääkkönen P, Ovaska E. Evaluating the quality of social media data in big data architecture[J]. IEEE Access, 2015, 3: 2028-2043.
- [16]Haryadi A F, Hulstijn J, Wahyudi A, et al. Antecedents of big data quality: An empirical examination in financial service organizations[A]//Proceedings of 2016 IEEE International Conference on Big Data[C]. IEEE, 2016: 116-121.
- [17]王宏志.大数据质量管理:问题与研究进展[J].科技导报,2014,32(34):78-84.  
Wang Hongzhi. Big data quality management: Problems and progress[J]. Science & Technology Review, 2014, 32(34): 78-84. (in Chinese)
- [18]Juddoo S, George C, Duquenoy P, et al. Data governance in the health industry: Investigating data quality dimensions

- within a big data context[J]. *Applied System Innovation*, 2018, 1(4): 1-16+43.
- [19] Firmani D, Mecella M, Scannapieco M, et al. On the meaningfulness of “Big Data Quality” [J]. *Data Science & Engineering*, 2016, 1(1): 6-20.
- [20] 王力, 周晓剑. 大数据质量评估的标准及过程研究[J]. *经营与管理*, 2018, (4): 84-88.  
Wang Li, Zhou Xiaojian. Research on standards and processes of big data quality assessment[J]. *Management and Administration*, 2018, (4): 84-88. (in Chinese)
- [21] 潘云鹤. AI及机器人的新方向[J]. *机器人技术与应用*, 2019, (4): 19-20.  
Pan Yunhe. New directions for AI and robotics[J]. *Robot Technique and Application*, 2019, (4): 19-20. (in Chinese)
- [22] 李国杰, 徐志伟. 从信息技术的发展态势看新经济[J]. *中国科学院院刊*, 2017, 32(3): 233-238.  
Li Guojie, Xu Zhiwei. Judging new economy from perspective of information technology trend[J]. *Bulletin of Chinese Academy of Sciences*, 2017, 32(3): 233-238. (in Chinese)
- [23] 顾基发, 唐锡晋, 朱正祥. 物理-事理-人理系统方法论综述[J]. *交通运输系统工程与信息*, 2007, (6): 51-60.  
Gu Jifa, Tang Xijin, Zhu Zhengxiang. Review of Wuli-Shili-Renli system approach[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2007, (6): 51-60. (in Chinese)
- [24] 黄婕, 安小米, 许济沧, 等. 基于国际标准的“数据利用”核心概念及概念体系研究[J]. *图书情报知识*, 2021, 38(5): 48-62.  
Huang Jie, An Xiaomi, Xu Jicang, et al. Research on core concepts and concept system of “Data Use” based on international standards[J]. *Documentation, Information & Knowledge*, 2021, 38(5): 48-62. (in Chinese)
- [25] 安小米, 王丽丽, 许济沧, 等. 我国政府数据治理与利用能力框架构建研究[J]. *图书情报知识*, 2021, 38(5): 34-47.  
An Xiaomi, Wang Lili, Xu Jicang, et al. Research on the framework construction of government data governance and data use capability in China[J]. *Documentation, Information & Knowledge*, 2021, 38(5): 34-47. (in Chinese)
- [26] 钱学森, 于景元, 戴汝为. 一个科学新领域——开放的复杂巨系统及其方法论[J]. *自然杂志*, 1990, (1): 3-10+64.  
Qian Xuesen, Yu Jingyuan, Dai Ruwei. A new field of science: Open complex giant system and its methodology[J]. *Chinese Journal of Nature*, 1990, (1): 3-10+64. (in Chinese)
- [27] 安小米, 马广惠, 宋刚. 综合集成方法研究的起源及其演进发展[J]. *系统工程*, 2018, 36(10): 1-13.  
An Xiaomi, Ma Guanghui, Song Gang. Origins and evolution of meta-synthesis approach[J]. *Systems Engineering*, 2018, 36(10): 1-13. (in Chinese)
- [28] 全国信息技术标准化技术委员会. 数据管理能力成熟度评估模型: GB/T 36073-2018[S]. 北京: 中国标准出版社, 2018: 24-28.  
China National Information Technology Standardization Technical Committee. Data management capability maturity assessment model: GB/T 36073-2018[S]. Beijing: Standards Press of China, 2018: 24-28. (in Chinese)
- [29] Data quality: Part 63: Data quality management; Process measurement; ISO 8000-63:2019[S/OL]. [2021-10-21]. <https://www.iso.org/standard/65344.html>.
- [30] 全国信息技术标准化技术委员会. 信息技术 大数据 大数据系统基本要求: GB/T 38673-2020[S]. 北京: 中国标准出版社, 2020: 6-8.  
China National Information Technology Standardization Technical Committee. Information technology-Big data-Basic requirements for big data systems: GB/T 38673-2020[S]. Beijing: Standards Press of China, 2020: 6-8. (in Chinese)
- [31] Information technology: Governance of IT for the organization; ISO/IEC 38500: 2015[S/OL]. [2021-10-21]. <https://www.iso.org/standard/62816.html>.
- [32] Systems and software engineering; Eystems and software quality requirements and evaluation (SQuaRE); Measurement of data quality; ISO/IEC 25024:2015[S/OL]. [2021-10-21]. <https://www.iso.org/standard/35749.html>.



- [33] 马一鸣. 政府大数据质量评价体系构建研究[D]. 长春: 吉林大学, 2016.  
Ma Yiming. The evaluation system construction of government big data quality[D]. Changchun: Jilin University, 2016. (in Chinese)
- [34] 洪学海, 王志强, 杨青海. 面向共享的政府大数据质量标准化问题研究[J]. 大数据, 2017, 3(3): 44 – 52.  
Hong Xuehai, Wang Zhiqiang, Yang Qinghai. Research on the quality control of sharing big data for government[J]. Big Data Research, 2017, 3(3): 44 – 52. (in Chinese)
- [35] 程 豪. 何为大数据质量评估框架——*Measurement Data Quality for Ongoing Improvement*[J]. 中国统计, 2016, (6): 18 – 19.  
Cheng Hao. Big data quality assessment framework: *Measurement Data Quality for Ongoing Improvement*[J]. China Statistics, 2016, (6): 18 – 19. (in Chinese)
- [36] 仲苏阳. 基于 MMTD 的大数据质量评价方法研究[D]. 南京: 南京邮电大学, 2017.  
Zhong Suyang. Research on big data quality evaluation based on MMTD[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2017. (in Chinese)
- [37] 宋朋波, 刘伊生, 郑 旺. 基于 ISM 的公共建筑大数据质量影响因素研究[J]. 河南科学, 2021, 39(6): 1025 – 1032.  
Song Pengbo, Liu Yisheng, Zheng Wang. Research on influencing factors of big data quality of public buildings based on interpretive structural modeling[J]. Henan Science, 2021, 39(6): 1025 – 1032. (in Chinese)
- [38] 张 坦, 黄 伟, 石 勇. ISO 8000(大)数据质量标准及应用[J]. 大数据, 2017, 3(1): 3 – 11.  
Zhang Tan, Huang Wei, Shi Yong. ISO 8000 (big) data quality standard and application[J]. Big Data Research, 2017, 3(1): 3 – 11. (in Chinese)
- [39] 周艳红. 电商大数据质量评价模型的建立及实证研究[D]. 重庆: 重庆工商大学, 2021.  
Zhou Yanhong. Establishment and Empirical Research of Ecommerce Big Data Quality Evaluation Model[D]. Chongqing: Chongqing Technology and Business University, 2021. (in Chinese)
- [40] 任照博. 航空经济大数据质量评价指标体系研究[D]. 郑州: 郑州航空工业管理学院, 2019.  
Ren Zhaobo. Research on Quality Evaluation Indicator System of Big Data in Aviation Economy[D]. Zhengzhou: Zhengzhou University of Aeronautics, 2019. (in Chinese)
- [41] 熊兴江. 医疗大数据质量评价指标体系构建研究[D]. 武汉: 华中科技大学, 2019.  
Xiong Xingjiang. Research on the Establishment of the Data Quality Evaluation Indicator System for Healthcare Big Data [D]. Wuhan: Huazhong University of Science & Technology, 2019. (in Chinese)
- [42] 漆 源, 王非函, 高洪美, 等. 基于层次分析法的公共安全大数据质量评估研究[J]. 现代信息技术, 2019, 3(3): 139 – 141 + 144.  
Qi Yuan, Wang Feihan, Gao Hongmei, et al. Research on data quality assessment for big data in public safety based on analytic hierarchy process[J]. Modern Information Technology, 2019, 3(3): 139 – 141 + 144. (in Chinese)
- [43] 赵 冰, 李 平, 代明睿. 铁路大数据质量评估与优化方法研究[J]. 中国铁路, 2018, (2): 63 – 67.  
Zhao Bing, Li Ping, Dai Mingrui. Methods for evaluating and improving quality of big data of railway[J]. China Railway, 2018, (2): 63 – 67. (in Chinese)
- [44] El Alaoui I, Gahi Y, Messoussi R. Big data quality metrics for sentiment analysis approaches[C]//ACM International Conference Proceeding Series, Hong Kong, 2019: 36 – 43.
- [45] Wahyudi A, Kuk G, Janssen M. A process pattern model for tackling and improving big data quality[J]. Information Systems Frontiers, 2018, 20(3): 467 – 469.
- [46] Abdallah M. Big data quality challenges[C]. Proceedings of the 2019 International Conference on Big Data and Computational Intelligence (ICBDCl), Pointe Aux Piments, 2019: 8686099.
- [47] Data quality: Part 8: Information and data quality: Concepts and measuring; ISO 8000 – 8:2015[S/OL]. [2021 – 10 – 21]. <https://www.iso.org/standard/60805.html>.

- [48] Data quality: Part 61: Data quality management: Process reference model: ISO 8000-61: 2016[S/OL]. [2021-10-21]. <https://www.iso.org/standard/63086.html>.
- [49] Terminology Work: Principles and Methods: ISO 704: 2022[S/OL]. [2021-09-20]. <https://www.iso.org/standard/38109.html>.
- [50] 罗伯特 K 殷. 案例研究方法的应用[M]. 重庆: 重庆大学出版社, 2004.  
Yin R K. Application of Case Study Method[M]. Chongqing: Chongqing University Press, 2004. (in Chinese)

## Construction of panoramic big data quality evaluation indicator framework

AN Xiao-mi<sup>1, 2, 3</sup>, HUANG Jie<sup>1, 3\*</sup>, XU Ji-cang<sup>1, 3</sup>, WANG Li-li<sup>1, 3</sup>, HONG Xue-hai<sup>4, 5</sup>,  
WANG Zhi-qiang<sup>6</sup>, HAN Xin-yi<sup>1</sup>

1. School of Information Resource Management, Renmin University of China, Beijing 100872, China;
2. Key Laboratory of Data Engineering and Knowledge Engineering, Ministry of Education, Beijing 100872, China;
3. Smart City Research Center, Renmin University of China, Beijing 100872, China;
4. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100086, China;
5. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100086, China;
6. China National Institute of Standardization, Beijing 100088, China

**Abstract:** Big data quality evaluation is an important guarantee for the high-quality development of digital economy, digital society and digital government. In view of the lacking of guidance of standardized document source and panoramic evaluation for current big data quality evaluation indicators, this paper summarizes the scenarios of big data quality evaluation application. Guided by Meta-synthesis, this paper proposes a panoramic big data quality evaluation perspective composed of “Human-Business-Data-Artifact”(HBDA). Content analysis, coding and case study methods are employed. Based on the representative literatures, a comprehensive big data quality evaluation indicator framework consisting of 56 indicators is constructed after two rounds of screening by three researchers. Then, big data quality evaluations are discussed in three practical case studies: B-City International Data Exchange, B-City Municipal Bureau of Coordinated Administrative Law Enforcement for Urban Management and B-City Big Data center. The case study result effectively supports the correctness and usability of the framework. The proposed indicator framework extends the application of the panoramic PAGE framework in multiple scenarios of big data quality evaluation and innovates the multi-dimensional standardized collaborative path of big data quality evaluation. It is of strategic significance to the overall improvement of big data quality in the construction of digital economy, digital society and digital government, and it is of guiding significance to enhance the data governance capacity, big data-driven management and decision-making capacity of digital China.

**Key words:** big data quality; evaluation indicators; framework construction; panoramic framework; the perspective of HBDA